		...cs	Search Text	DBs	Time Stamp	Comments	Error Definition	Errors
38	BRS	4	user near3 affinity near5 image	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/08 16:59			0
39	IS&R	2	("6317722").PN.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 09:30			0
40	BRS	11	("4870579" | "4996642" | "5235509" | "5459306" | "5583763" | "5745681" | "5749081" | "5774670" | "5790935" | "5905973" | "5909492").PN.	USPAT	2003/09/08 16:59			0
41	BRS	4338	significan$5 with affinity	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 09:30			0
42	BRS	1	(significan$5 with affinity) with entity	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 09:31			0
43	BRS	20	(significan$5 with affinity) same entity	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 09:31			0
44	BRS	30	(significan$5 with affinity) same itera$8	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 09:32			0
45	BRS	30	((significan$5 with affinity) same itera$8) and similar$5	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 09:38			0
46	BRS	157	(significan$5 with affinity) with similar$5	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 09:38			0
47	BRS	11	((significan$5 with affinity) with similar$5) and itera$8	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 09:40			0
48	BRS	2	affinity near8 significance near8 value	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 09:41			0
49	BRS	351	"affinity value"	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 09:41			0
50	BRS	527	"significance value"	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 09:41			0
51	BRS	0	"affinity value" same "significance value"	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 09:41			0
52	BRS	0	"affinity value" and "significance value"	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 09:42			0
53	BRS	112	affinity near8 significance	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 09:42			0
54	BRS	10	(affinity near8 significance) and iterat$8	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 09:49			0

	Type	Hits	Search Text	DBs	Time Stamp	Comments	Error Definition	Errors
55	IS&R	2	("6389436").PN.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 10:26			0
56	BRS	801	"common subject matter"	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 10:27			0
57	BRS	0	"common subject matter" near8 significan$5	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 10:27			0
58	BRS	0	"common subject matter" same significan$5 same affinity	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 10:27			0
59	BRS	10	"common subject matter" and significan$5 and affinity	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 10:28			0
60	BRS	17	"common subject matter" same link	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 10:45			0
61	BRS	3	"6112202" and affinity	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 10:45			0
62	BRS	14	"6112202"	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 10:47			0
63	IS&R	2	("6112202").PN.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/09/09 10:47			0

US006356899B1

(12) **United States Patent**
Chakrabarti et al.

(10) **Patent No.: US 6,356,899 B1**
(45) **Date of Patent: *Mar. 12, 2002**

(54) **METHOD FOR INTERACTIVELY CREATING AN INFORMATION DATABASE INCLUDING PREFERRED INFORMATION ELEMENTS, SUCH AS PREFERRED-AUTHORITY, WORLD WIDE WEB PAGES**

(75) Inventors: **Soumen Chakrabarti**, Maharashtra (IN); **Byron Edward Dom**, Los Gatos, CA (US); **David Andrew Gibson**, Berkeley, CA (US); **Prabhakar Raghavan**, Saratoga, CA (US); **Sridhar Rajagopalan**, San Jose, CA (US); **Shanmugasundaram Ravikumar**, San Jose, CA (US); **Andrew Tomkins**, San Jose, CA (US)

(73) Assignee: **International Business Machines Corporation**, Armonk, NY (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **09/261,926**

(22) Filed: **Mar. 3, 1999**

Related U.S. Application Data

(63) Continuation-in-part of application No. 09/143,733, filed on Aug. 29, 1998.

(51) **Int. Cl.**[7] **G06F 17/30**
(52) **U.S. Cl.** **707/5**; 707/3; 707/4; 707/104
(58) **Field of Search** 707/1, 3, 4, 5, 707/104; 709/218, 219

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,870,579 A 9/1989 Hey 705/27
4,945,475 A 7/1990 Bruffey et al. 707/1
4,996,642 A 2/1991 Hey 705/27

(List continued on next page.)

OTHER PUBLICATIONS

Indexing for linear model–based information retrieval by—Yuan–Chi Chang; Chung–Sheng Li; IBM Thomas J. Watson Res. Center, Yorktown Heights, NY, USA ;in: Multimedia and Expo, 2000. ICME 2000. 2000 IEEE; International Conference; pp: 359–362.*

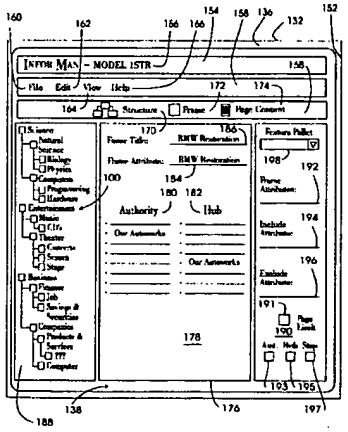
(List continued on next page.)

Primary Examiner—Diane D. Mizrahi
(74) *Attorney, Agent, or Firm*—John L. Rogitz

(57) **ABSTRACT**

A method for identifying, filtering, ranking and cataloging information elements; as for example, World Wide Web pages, of the Internet in whole, part, or in combination. The method is preferably implemented in computer software and features steps for enabling a user to interactively create an information database including preferred information elements such as preferred World Wide Web pages in whole, part, or in combination. The method includes steps for enabling a user to interactively create a frame-based, hierarchical organizational structure for the information elements, and steps for identifying and automatically filtering and ranking by relevance, information elements, such as World Wide Web pages for populating the structure, to form; for example, a searchable, World Wide Web page database. Additionally, the method features steps for enabling a user to interactively define a frame-based, hierarchical information structure for cataloging information, identifying a preliminary population of information elements for a particular hierarchical category arranged as a frame, based upon the respective frame attributes, and thereafter, expanding the information population to include related information, and subsequently, automatically filtering and ranking the information based upon relevance, and then populating the hierarchical structure with the a definable portion of the filtered, ranked information elements.

50 Claims, 19 Drawing Sheets



U.S. PATENT DOCUMENTS

5,230,072	A		7/1993	Smith et al. 707/1
5,535,382	A	*	7/1996	Ogawa 707/5
5,583,763	A		12/1996	Atcheson et al. 707/3
5,615,341	A		3/1997	Agrawal et al. 705/10
5,845,270	A		12/1998	Schatz et al. 706/11
5,940,821	A	*	8/1999	Wical 707/3
5,987,454	A	*	11/1999	Hobbs 707/4

OTHER PUBLICATIONS

Microsoft Office 6–in–1 New Edition(1994, pp. 1–13 by Que Corporations.*

Publication: "Inferring Web Communities from Link Topology." Gibson et al. Proceedings of the 9th ACM Conference on Hypertext and Hypermedia. pp 1–17. 1998.

Publication: "Automatic Resource Compilation by Analyzing Hyperlink Structure and Associated Text." Chakrabarti et al. Proceedings of the 7th World Wide Web Conference. 1998.

Publication: "Visualizing Rich, Structured Hypermedia". Keith Andrews, Information Visualization Short Note, Graz University of Technology, pp. 40. Jul., 1998.

Brochure: Microsoft Office 6–in–1, New Edition. Jul., 1994. (p. 1).

Publication: "Authoritative Sources in a Hyperlinked Environment". Kleinberg. Proceedings of the ACM–SIAM Symposium on Discrete Algorithms, pp. 1–31. 1998.

Publication: "Beehive: A System for Cooperative Filtering and Sharing of Information". Huberman et al. Dynamics of Computation Group, Xerox Palo Alto Research Center, Palo Alto, CA. pp. 1–9. Aug. 1996.

Publication: "Using Collaborative Filtering to Weave an Information Tapestry". Goldberg et al. Communication of the ACM. vol. 35, No. 12. Dec., 1992.

Publication: "The Strength of Weak Ties". Mark S. Granovetter. The American Journal of Sociology. vol. 78, No. 6. pp. 1360–1380, 1973.

Publication: "GroupLens: An Open Architecture for Collaborative Filtering of Netnews." Resnick et al. ACM. pp. 175–186. 1994.

Publication: "Knowledge Integration for Structured Information Sources Containing Text (Extended Abstract)". Cohen. SIGIR–97 Workshop on Networked Information Retrieval. pp. 0–17. Aug., 1997.

Publication: "Do–I–Care: A Collaborative Web Agent." Starr et al. CHI '96 Companion. pp. 273–274. Apr. 1996.

Publication: "Clustering and Information Sharing in an Ecology of Cooperating Agents." Leonard N. Foner. 1995.

Publication: "Evolving Agents For Personalized Information Filtering." Sheth et al. IEEE. pp. 345–352. 1993.

Publication: "Experiences with GroupLens: Making Usenet Useful Again." Miller et al. Annual Technical Conference. pp. 219–233. 1997.

Publication: "Recommender Systems." Resnick et al. Communications of the ACM. vol. 40., No.3, pp. 56–89. Mar., 1997.

Publication: "Recommending and Evaluating Choices in a Virtual Community of Use." Hill et al. CHI '95 Mosaic of Creativity, pp. 194–201. May, 1995.

Publication: "Pointing The Way: Active Collaborative Filtering." Maltz et al. CHI '95 Mosaic of Creativity, pp. 202–209. May, 1995.

Publication: Social Information Filtering: Algorithms for Automating "Word of Mouth." Shardanand et al. CHI '95 Mosaic of Creativity, pp. 210–217. May, 1995.

* cited by examiner

US006112202A

United States Patent [19]

Kleinberg

[11] Patent Number: 6,112,202

[45] Date of Patent: Aug. 29, 2000

[54] **METHOD AND SYSTEM FOR IDENTIFYING AUTHORITATIVE INFORMATION RESOURCES IN AN ENVIRONMENT WITH CONTENT-BASED LINKS BETWEEN INFORMATION RESOURCES**

[75] Inventor: **Jon Michael Kleinberg**, Los Gatos, Calif.

[73] Assignee: **International Business Machines Corporation**, Armonk, N.Y.

[21] Appl. No.: **08/813,749**

[22] Filed: **Mar. 7, 1997**

[51] **Int. Cl.**[7] **G06F 17/30**

[52] **U.S. Cl.** **707/5**; 707/9; 707/101

[58] **Field of Search** 707/1, 2, 4, 5, 707/10, 100, 101, 102, 501

[56] **References Cited**

U.S. PATENT DOCUMENTS

5,257,185	10/1993	Farley et al.	707/100
5,446,891	8/1995	Kaplan et al.	707/2
5,778,363	7/1998	Light	707/5
5,826,031	10/1998	Nielsen	395/200.63
5,835,905	11/1998	Pirolli et al.	707/3

OTHER PUBLICATIONS

Savoy, J., Searching Information in Hypertext Systmes using Multiple Sources of Evidence, International Journal of Man–Machine Studies, 1993, pp. 1017–1030.

Stieger, H., Making Use of Hypertext Links when Retrieving Information, ACM, pp. 102–111, 1992.

R.W. Schwanke et al., Cross References are Features, Sec. 10.1, Book/Machine Learning: From Theory to Applications, Cooperative Research at Siemens and MIT, Appeared in Proceedings of the 2nd International Workshop on Software Configuration Mng., Princeton, NJ, Oct. 1989, ACM SIGSoft, IEEE CS, and GI pp. 107–123.

H.C. Arents et al., "Concept–Based Retrieval of Hypermedia Information: From Term Indexing to Semantic Hyperindexing," Information processing & Management vol. 29, No. 3, pp. 373–386, 1993.

R. Rada et al., "Retrieval Hierarchies in Hypertext, " Information Processing & Mng., vol. 29, No. 3, (Printed in Great Britain) pp. 359–371, 1993.

W.M. Shaw, Jr., "Subject and Citation Indexing. Part 1: The Clustering Structure of Composite Representations in the Cystic Fibrosis Document Collection, " JASIS–Journal of the American Society for Information Science, vol. 42, No. 9, Oct. 1991, pp. 669–675.

W.M. Shaw, Jr., "Subject Indexing & Citation Indexing–Part II: A Evaluation and Comparison Information Processing & Management", vol. 26, No. 6, (printed in Great Britain) pp. 705–718, 1990.

T.R. Kochtanek, "Brief Communication, Document Clustering, Using Macro Retrieval Techniques, " Journal of the American Society for Information Science, vol. 34, No. 5, pp. 356–359, Sep. 1993.

F. Narin et al., Chapter 2., "Bibliometrics, " Pub. Annual Review of Information Science and Technology, pp. 35–58, 1977.

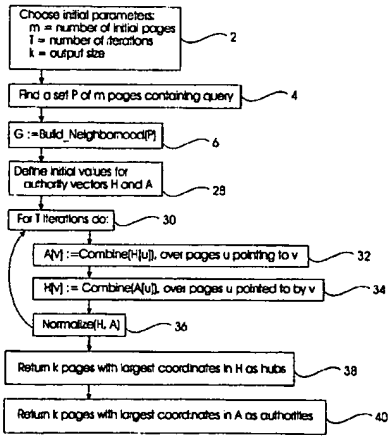
(List continued on next page.)

Primary Examiner—Thomas G. Black
Assistant Examiner—John Loomis
Attorney, Agent, or Firm—Khanh Q. Tran

[57] **ABSTRACT**

A system and method are provided for searching for desired items from a network of information resources. In particular, the system and method have advantageous applicability to searching for World Wide Web pages having desired content. An initial set of pages are selected, preferably by running a conventional keyword-based query, and then further selecting pages pointing to, or pointed to from, the pages found by the keyword-based query. Alternatively, the invention may be applied to a single page, where the initial set includes pages pointed to by the single page and pages which point to the single page. Then, iteratively, authoritativeness values are computed for the pages of the initial set, based on the number of links to and from the pages. One or more communities, or "neighborhoods", of related pages are defined based on the authoritativeness values thus produced. Such communities of pages are likely to be of particular interest and value to the user who is interested in the keyword-based query or the single page.

57 Claims, 5 Drawing Sheets



OTHER PUBLICATIONS

J. Bichteler et al, "Document Retrieval by Means of an Automatic Classification Algorithm for Citations, " Information Storage Retr. vol. 10, pp. 267–278, (Printed in Great Britain), 1974.

W.M. Shaw, Jr., "Subject and Citation Indexing Part II: Optimal, Cluster–Based Retrieval Performance of Composite Representations", Journal of the American Society for Information Science, vol. 42, No. 9, pp. 676–684, Oct. 1991.

M.E. Frisse, "Searching for Information in a Hypertext Medical Handbook, " Communications of the ACM, vol. 31, No. 7, pp. 880–886, Jul. 1988.

E. Rivlin., "Navigating in Hyperspace: Designing a Structure–Based Toolbox, " Navigating in Hyperspace, Comm. of ACM, vol. 37, No. 2, pp. 87–96, Feb. 1994.

R. Weiss et al., HyPursuit: A Hierarchical Network Search Engine that Exploits Content–Link Hypertext Clustering, Programming Systems Research Group, MIT Lab. For Computer Sci., 545 Technology Square, Cambridge, MA 02139, no date.

D.A. Speilman et al., Spectral Partitioning Works: Planar graphs and finite element meshes., Abstract based on UC Berkeley Tech. Report, UCB/CSD–96–898, no date.

G. H. Golub, (book) Matrix Computations, 2nd Edit., Ch. 5, Orthogonalization & Least Squares, 5.2.7 Classical Gram–Schmidt, pp. 218–219; Ch. 7 The Unsymmetric Eigenvalue Problem, 7.3 Power Iterations, pp. 351–354, Orig. Pub. 1989.